









#### 4.2 Replacing BERT to other models

除了對比人為設置的學習率參數以及損失函數，本實作還對採用的語言模型做了對比實驗。

本實作所對比之模型主要為目前主流之語言預訓練模型，RoBERTa(Liu et al. 2019)、BART(Lewis et al. 2020)。

實驗在 1.1、1.2、1.3 三個相對成果較好的學習率衰減率下以及 CRF 作為最後一層的設定下做了對比實驗，如表四所示。

其中不難看出，BERT-base 模型目前在該任務上能夠有更出色的效能，從 F1 分數上，能夠略強於其他兩種主流預訓練模型。也不難發現，從 F1 上看，BERT(Devlin et al., 2019)和 RoBERTa(Liu et al. 2019)的最佳衰減率都在 1.2 左右，而 BART 則在 1.3 左右。

### 5 Conclusion

本實作為 ROCLING 2022 SHARED TASK (Lee et al. 2022) 之生醫命名實體識別任務實作，參考了一些論文的實作方法，使用預訓練語言模型和一定程度的模型微調，來達到準確率局部最大化的目的。

通過對比實驗可知，CRF、學習率適當地逐層衰減以及 bert 預訓練模型在命名實體識別的任務上都能在一定程度上有所提升。

相比於模型微調的細節參數變化，系統效能與模型結構和預訓練方法對結果的影響可能更加具有決定作用，因此本實作若可能，之後預期在模型結構、預訓練等方面進行進一步細節上的處理，以提高效能。

### 6 Reference

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, “Attention Is All You Need” in NeurIPS 2017

Lung-Hao Lee, Chao-Yi Chen, Liang-Chih Yu, and Yuen-Hsien Tseng. 2022. Overview of the ROCLING 2022 shared task for Chinese healthcare named entity recognition. In Proceedings of the 34th Conference on Computational Linguistics and Speech Processing.

Lung-Hao Lee and Yi Lu, "Multiple Embeddings Enhanced Multi-Graph Neural Networks for Chinese Healthcare Named Entity Recognition," in IEEE Journal of Biomedical and Health Informatics, 2021, pp. 2801-2810

Jeremy Howard, Sebastian Ruder, “Universal Language Model Fine-tuning for Text Classification” in ACL 2018, pp. 328-339

Xuezhe Ma and Eduard Hovy, “End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF” in ACL 2016, pp. 1064–1074

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, Chris Dyer, “Neural Architectures for Named Entity Recognition” in NAACL 2016, pp. 260–270

Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” in NAACL 2019, N19-1423, pp. 4171–4186

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdel-rahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer, “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension” in 2020.acl-main.703, pp. 7871–7880

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov, 2019, “RoBERTa: A Robustly Optimized BERT Pretraining Approach”, arXiv:1907.11692

Sascha Rothe, Shashi Narayan, Aliaksei Severyn, TACL 2020, “Leveraging Pre-trained Checkpoints for Sequence Generation Tasks” in 2020.tacl-1.18, pp. 264–280

Zhiheng Huang, Wei Xu, Kai Yu, “Bidirectional LSTM-CRF Models For Sequence Tagging”, arXiv:1508.01991, 2015

Shikhar Vashishth, Shyam Upadhyay, Gaurav Singh Tomar, Manaal Faruqui, “Attention Interpretability Across NLP Tasks”, 2019, arXiv:1909.11218

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, “Improving Language Understanding by Generative Pre-Training” in OpenAI, 2018

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, Veselin Stoyanov, “Unsupervised Cross-lingual Representation Learning at Scale” in ACL 2020, 2020.acl-main.747, pp. 8440–8451